

Introduction/Business Problem

Identifying key locations of where accidents most frequently occur as well as any possible factors involved with those locations may help mitigate city/county costs associated with automobile and pedestrian accidents. Once identified, locations and factors can be addressed with infrastructure upgrades to further reduce occurrence and severity of accidents, reduce legal costs for the city and taxpayer, and reduce the human cost associated with accidents. For instance, identifying locations where weather (notably rain) is a clear factor in accident severity will allow city/state resources to improve infrastructure to decrease accident occurrences. Similar cases can be made to factors such as light conditions or types of accident. City and state taxpayers would benefit from this analysis through reduced long-term repair, and legal costs associated with accidents, as well as improve taxpayer satisfaction with allocation of city/state resources.

To summarize, we wish to solve the longstanding problem of high accident occurrence and severity as a function of location and conditions. Our audience are city/state taxpayers, contractors, and planners, who directly benefit from the improvement of city/state infrastructure by reducing legal and human costs associated with accidents.

Data

The data is obtained from the Seattle Department of Transportation and consists of collision data for all years. The target labels for this dataset are *accident severity*, measured on a scale from 0 to 3, with 0 being the least severe and 3 being the most severe. In total, there are 37 columns, however, some of these attributes, notably the text descriptions, will not be useful in the classification model. The attributes (predictors) of the dataset, which can be used in the supervised classification model, include positional (latitude and longitude) coordinates, road conditions, collision types, and the number of vehicles, pedestrians, and injuries associated with each collision. The model can then be used to identify locations and conditions where accidents are the most severe. The results of the model can allow city/state contractors and engineers to target infrastructure that requires upgrades to decrease the occurrence or severity of accidents in the future.